www.ierjournal.org

ISSN 2395-1621



Detection of Suicidal Ideation from Twitter Post Using Machine Learning

Paritosh Sham Pardeshi, Priyanka Vilas Dongre, Sneha Sunil Deokar, Snehal Bharat Ahire, Prof. Satish N Gujar

Department of Computer Engineering

TSSM BSCOER, NARHE, PUNE

ABSTRACT

Suicidal ideation is one of the most severe mental health issues faced by people all over the world. There are various risk factors involved that can lead to suicide. The most common critical risk factors among them are depression, anxiety, social isolation and hopelessness. Early detection of these risk factors can help in preventing or reducing the number of suicides. Online social networking platforms like Twitte, Redditt and Facebook are becoming a new way for the people to express themselves freely without worrying about social stigma. This paper presents a methodology and experimentation using social media as a tool to analyse the suicidal ideation in a better way, thus helping in preventing the chances of being the victim of this unfortunate mental disorder. The data is collected from Twitter, one of the popular Social Networking Sites (SNS). The Tweets are then pre-processed and annotated manually. Finally, various machine learning and ensemble methods are used to automatically distinguish Suicidal and Non-Suicidal tweets. This experimental study will help the researchers to know and understand how SNS are used by the people to express their distress related feelings and emotions. The study further confirmed that it is possible to analyse and differentiate these tweets using human coding and then replicate the accuracy by machine classification. However, the power of prediction for detecting genuine suicidality is not confirmed yet, and this study does not directly communicate and intervene the people having suicidal behaviour.

ARTICLE INFO

Article History

Received: 15th April 2022 Received in revised form : 15th April 2022 Accepted: 20th April 2022 **Published online :** 21st April 2022

Keywords: Suicidal, Social Media, Twitter Dataset, Random Forest, Naive Bayes, KNN, etc.

I. INTRODUCTION

Suicide is one of the significant public health concerns consuming a lot of lives. According to the statistic of the World Health Organisation (WHO) (1), around one million people die due to suicide each year, and on average, suicide occurs every 40 seconds. Among the total suicide-related deaths, 135000 deaths occurred in India alone (2). WHO further mentioned that suicide is the primary cause of death among teenagers and the sixth leading cause among adults. Furthermore, there are 20 times more suicidal attempts disrupting the families emotionally and economically.

American Foundation for Suicide Prevention (AFSP) has identified various risk factors associated with suicide. The factors include personal issues like hopelessness, substance abuse, anxiety, schizophrenia; social factors like isolation from society, loss of loved ones, unemployment, bullying or abuse, or factors related to negative events in life like illness, emotional disorders, and history of previous suicide attempts. It is a common saying that suicide is a permanent solution for dealing with temporary problems. Despite the growing numbers of suicidal cases, it can be prevented to some extent by understanding the risk factors related to suicidal behaviour in the early stages of the suicidal process.

The process of suicide starts with suicidal thoughts or ideation. It then matures to suicidal attempt and finally to the completed suicide. Prevention can be done by reducing the risk factors or by reducing the obstacles to mental health resources.

II. PROBLEM STATEMENT

Suicidal ideation is one of the most severe mental health issues faced by people all over the world. There are various risk factors involved that can lead to suicide.

The most common & critical risk factors among them are depression, anxiety, social isolation and hopelessness.

Early detection of these risk factors can help in preventing or reducing the number of suicides. Online social networking platforms like Twitter, Redditt and Facebook are becoming a new way for the people to express themselves freely without worrying about social stigma.

Hence there is a need to develop such an application which will detect suicidal ideation.

To develop a web based prototype model using python for suicidal ideation on twitter using machine learning techniques.

III. LITERATURE SURVEY

Deep Learning for Hate Speech Detection in Tweets In this paper , the Author investigated the application of deep neural network architectures for the task of hate speech detection. The author found them to remarkably outperform the existing methods, which were able to classify a tweet as sexiest, racist, or neither. Embedded learning from deep neural network models when merged with gradient boosted decision trees led to best accuracy values. The author, performed extensive experiments to learn complexity handling of semantic word embedding's with multiple deep learning architectures. In the future, The Author plans to explore the significance of the user network features for the task.

2. Smokey: Automatic Recognition of Hostile Messages This paper, describes some approaches to flame recognition, including a prototype system, Smokey. Author combined natural-language processing and sociolinguistic observation for identifying messages containing insulting words used in an insulting manner. Smokey builds a 47- clement feature vector based on the syntax and the semantics of every sentence, combining the vectors for the sentences within each message. A training set of 720 messages was used by Quinlans C4.5 decision-tree generator to determine featurebased rules that Author's ability to correctly sort 64 percent of the flames and 98 percent of the no flames in another test set of 460 messages. Further techniques for greater accuracy and user customization are also discussed.

3. Design of a Scalable Data Stream Channel for Big Data Processing In this paper author, tried to outline big data infrastructure for processing a huge amount of data streams. The project is based on a distributed stream computing platform. It provides cost-effective and large-scale big data services by developing a data stream management system. The author focused on scalable data stream channel, big data processing, big data infrastructure, data streaming process, and management. This research contributes to advancing the feasibility of massive processing for distributed, real-time computation even once they are overloaded.

4. Developing a Real-time Data Analytics Framework for Twitter Streaming Data The proposed framework includes data ingestion, stream processing, and data visualization components with the Apache Kafka messaging system that is used to perform data ingestion tasks. Furthermore, Spark makes it feasible to perform sophisticated processing and machine learning algorithms in realtime. The key focus of the paper is on developing an analytical framework with the ability of in-memory processing to extract and analyze structured and unstructured Twitter data. It has even conducted a case study on tweets about the event of the Japan earthquake and how the people around the world reacted to it by analyzing the time and origin of tweets.

5. Detection of Behavior Patterns through Social Networks like Twitter, using Data Mining techniques as a method to detect Cyber bullying This research focuses on the detection and analysis of cyber bullying on pages and with pejorative terms in Spanish, taking profit of the author of classification of feelings through specialized tools. For the detection of cyberbullying, first, the efficiency of classification of every tool is measured, through a group of pejorative terms commonly wont to hurt people. In the analysis stage author used data mining techniques to generate a dictionary of pejorative terms that are related to cyber bullying and was thus be able to generate behavior patterns of these terms. This way provided better tools so that psychology specialists can optimize their work.





Fig. 1 System architecture

www.ierjournal.org

Front End: Front End will be developed using html technology.

Back End: Back will be developed using Python and OpenCV. Back end contains preprocessing, feature extraction and classification.

Dataset: Here we are going to use twitter dataset for detection of suicidal tweets for which public tweets will be downloaded from twitter, with hash tag suicidal and a training model will be build using random forest algorithm.

V. ALGORITHM USED

Random Forest:

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

VI. CONCLUSION

Thus we are going to develop a web based prototype model for twitter where tweets will be classified as suicidal or normal.

If tweet will be identified with suicidal ideation, automatically a mail will be sent to the authority which will take care further.

This project if implemented by SNS, can save someone's life.

VII.REFERENCES

[1] Desmet B, Hoste V. Online suicide prevention through optimised text classification. Inf Sci (Ny). 2018;439–440:61–78.

[2] Patel V, Ramasundarahettige C, Vijayakumar L, Thakur JS, Gajalakshmi V, Gururaj G, et al. Suicide mortality in India: A nationally representative survey. Vol. 379, The Lancet. Lancet Publishing Group; p. 2343–51.

[3] Fu K, Cheng Q, Wong PWC, Yip PSF. Responses to a Self-Presented Suicide Attempt in Social Media. Crisis.;34(6):406–12.

[4] Gunn JF, Lester D. Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death. Suicidologi. 2015 Jun 1;17(3).

[5] Kailasam VK, Samuels E. Can social media help mental health practitioners prevent suicides? Curr Psychiatr.):37–39, 51.